

sustainable electronic records

STRATEGIES FOR THE MAINTENANCE AND
PRESERVATION OF ELECTRONIC RECORDS AND
DOCUMENTS IN THE TRANSITION TO 2004

VERSION 1.0

AUGUST 2001

© Crown copyright 2001

First version published 2001

Public Record Office
Ruskin Avenue
Kew
Surrey
TW9 4DU

email: e-records@pro.gov.uk

Website: www.pro.gov.uk/recordsmanagement/

Designed and typeset by Format Information Design

www.format-info.co.uk

Contents

1	Summary	1
2	Introduction	3
	Aims and intended use	
	Audience	
3	Benefits: why sustainability is important	7
	Sustaining the organisational business need	
	Sustaining legislative obligations	
4	Key elements for a 2004 compliant sustainability plan	9
	Rationale and essential prerequisites	
	Issues	
	Criteria for an effective plan	
	Quality indicators to assess effectiveness of the plan	
	Prioritising the work – use of plans and timetables	
	Management of relations with IT providers	
	Business contingency/continuity planning	
	Legal admissibility and evidential weight	
5	Access	15
	Who needs access	
	How access is provided	
	Frequency of access	
	Disaster recovery	
	Vital and emergency records	
	Checklist	
6	Options	18
	Summary	
	Storage of documents in native format on existing platforms	
	Migration strategies and emulation environments	
	Migration to standardised formats on existing platforms	
	Storage in native format on managed EDRMS environments	
	Storage in standardised format on managed EDRMS environments	
	Best practice management of databases on existing platforms	
	Classifying a database	
	Determining what to capture	
	Records no longer required for business purposes	

1 Summary

1.1 This toolkit is provided to help departments establish a strategic plan to ensure the records listed in departmental inventories of electronic record collections are adequately maintained and sustained whilst a residual need for these records continues to exist. The responsible post-holders will vary according to the organisation but it is anticipated that this document will assist those managers charged with the implementation of the Modernising Government agenda, which requires all newly created public records to be electronically stored and retrieved by 2004. This requires departments to undertake appropriate action to meet specific milestones at designated dates. One milestone requires the development of a sustainability strategy for maintaining access to, and reliability of, electronic documents identified as having continuing value by September 2001.

1.2 Departments were required to develop an up-to-date and comprehensive inventory of records collections covering any electronic records or electronic information which is regarded as meriting treatment as an electronic record by the end of December 2000. Departments then needed to establish both an evaluation plan to examine the contents of these inventories to determine value and correct disposal, and a preservation strategy in order to define the requirements to sustain the records and to prioritise action on:

- applying a departmental preservation or sustainability strategy to them as a part of that managed environment.

1.3 Electronic records that organisations require to be retained for business purposes need to be brought into a managed environment to sustain access for as long as they are required. Departments should now be planning to evaluate the electronic record collections listed in their inventories in order identify which records require maintenance for future access.

1.4 Departments should also be developing an accompanying sustainability strategy for maintaining access to, and reliability of, electronic documents identified as having continuing value by September 2001. This in turn will enable departments to develop a process for migrating records to new IT platforms as necessary including migration to a future corporate ERM system.

1.5 This toolkit has been produced to explain what now needs to be done to meet the target to develop an appropriate strategy to maintain these records for as long as they are needed. In order to comply with the milestone departments need to:

- plan to bring those electronic records that are valuable and still required for business and other reasons, into a managed environment, and then
- apply a departmental digital preservation strategy to them as a part of that managed environment.

1.6 This toolkit will help departments develop three main products:

- develop a compliant sustainability plan
- identify the access requirements for those records that are to be maintained
- identify the appropriate options needed to maintain the records over time.

1.7 For documents and records which are deemed to have continuing value, there are four options:

- maintain in native format on existing technical platform
- maintain in a standard format on existing technical platform
- migrate in native format to new ERM systems
- migrate in a standard format to new ERM systems

1.8 The toolkit also includes some summary advice on the maintenance of structured databases and datasets (to be expanded in a later toolkit).

1.9 This toolkit (or the September 2001 milestone) does not attempt to address longer term issues of digital preservation, which will be necessary once corporate ERM is established. These are the subject of further work by the PRO, which will be released later in 2002.

2 Introduction

Aims and intended use

2.1 This toolkit is one of a set that provides practical guidance for public record bodies so that they are better placed to meet the 2004 target for electronic records within the Modernising Government strategy. This states that by 2004 all newly created public records will be electronically stored and retrieved. The toolkits support the work needed to meet the interim milestones, which central government departments are being asked to adhere to en route to this target. These milestones are the dates by which actions must to be completed to achieve the 2004 target.

2.2 The *e-Government Policy Framework for Electronic Records Management* provides a framework and a set of milestones for departments and agencies to move towards full electronic records management by: including ERM implications in e-business strategies; gaining formal control of existing electronic records that have value as evidence; and planning for corporate-wide implementation of electronic records management systems to meet the Modernising Government target – that by 2004 all newly created public records will be electronically stored and retrieved. It can be accessed on the Office of the e-Envoy's website at <http://www.e-envoy.gov.uk/publications/frameworks/erm2/execsumm.htm>

The *Route map and milestones to achieve electronic record management by 2004* gives further information on the transition and milestones towards 2004. It can be accessed on the PRO website at http://www.pro.gov.uk/recordsmanagement/eros/ROUTE_MAP_7.pdf

2.3 Departments were required to develop an up-to-date and comprehensive inventory of records collections covering any electronic records or potential electronic records by the end of December 2000. Building on this, departments then need to establish both an evaluation plan to examine the contents of these inventories to determine value and correct disposal, and a preservation strategy in order to define strategies to sustain the records prioritising action on:

- bringing those electronic records that are valuable and still required for business and other reasons, into a managed environment;
- applying a departmental preservation or sustainability strategy to them as a part of that managed environment.

2.4 The toolkit is intended as practical advice for departments who must – as part of their route to 2004 – begin to manage their currently unstructured and unmanaged electronic records and to ensure that records required to be retained for business purposes are sustained in a manner that both ensures access throughout the period of retention and appropriate audit information which would assist in the authentication of the record should that be necessary. The toolkit provides guidance on the necessary planning and strategic thinking required to develop appropriate strategies as opposed to detailed operational advice.

2.5 It is intended for use in the transition period to full electronic records management, and does not attempt to solve problems of longer-term digital preservation. The PRO is undertaking further work on this important area.

2.6 In order to ensure that records do not lose their value over time departments need to develop a strategy for maintaining access to needed records through changes in hardware and software, and for maintaining the accuracy and authenticity of records over time.

2.7 An inventory of records collections, together with an evaluation strategy, will enable an organisation to understand what electronic records exist, of what types and of what value, and to determine the actions and resource implications needed to maintain those with continuing value.

2.8 The preservation plan provides the mechanisms to ensure that those electronic records of continuing value are maintained so that they will not lose their value over time. This in turn feeds into the implementation of facilities and procedures for the management, control and preservation of all new electronic records, and will provide a platform for the continued development of a strategy for sustainable electronic records management.

2.9 This toolkit assumes that departments will evaluate existing electronic records to identify information flows, eliminate unnecessary duplication and determine appropriate retention periods. If existing electronic records are not evaluated, material in inventories will remain unmanaged and unsustainable and any preservation plan would be compromised.

2.10 This document identifies the issues to be addressed when considering the medium term preservation of electronic records, and provides criteria for the development of a strategic plan for the preservation of identified electronic records. It provides guidance on best practice but is not completely prescriptive. Departments should determine an approach which is most appropriate for their own circumstances. The document should be used in conjunction with the guidance in two other related publications *Guidance for an inventory of electronic record collections* and *Evaluating information assets: appraising the inventory of electronic records*.

2.11 Once the transition to full corporate ERM is in place – by 2004 – the inventory will no longer be required to manage existing material (except in the short-term for a small amount of residual material). Much existing material, which does not fall into this category, will have been migrated to an EDRM or other ERM system, or appropriately destroyed. The inventory may, however, provide the basis of a tool to continue management of structured databases and datasets. Additional strategies for the longer-term sustainable management of:

- corporate records in an EDRM system
- structured databases and datasets

will be required in the post-2004 landscape. The milestone for September 2001 is not intended to address these issues.

Audience

2.12 This toolkit is designed to help Departmental Record Officers and others charged with record management responsibilities to meet the 2004 target. There will be other stakeholders in the organisation who will participate in the assessment of electronic record collections, from an operational, business or IT perspective. All concerned should ensure consistency with the organisation's corporate policy and procedures, and general working practices.

2.13 The guidance is intended primarily for those working in central government; the principles will also be relevant in local government and throughout the public sector. Throughout this document the term 'department' should be taken to apply to any public sector organisation, including all departments, agencies and other organisations across government. Familiarity with the concepts of records as used in central government is assumed.

3 Benefits: why sustainability is important

Sustaining the organisational business need

3.1 The creating, owning and custodial organisation's intentions for preserving the records and their information content need to be clearly understood. Use of or reference to the records may well increase or diminish during the period of preservation as the statutory and administrative climate changes. As organisations start to develop and profit from their corporate information resources use of retained records may grow and they may be used in ways not foreseen by the creators. Consideration needs to be given to the development of appropriate mechanisms, where a business need can be foreseen, to ensure that these variations in demand are adequately addressed. In the sustainability plan, be clear about:

- why records are being preserved
- the terms which govern the period of preservation, including any changes
- the relevance to business needs

so that these factors are transparent to those individuals entrusted with the preservation of the records.

Sustaining legislative obligations

3.2 Data protection is subject to the overriding principle that the accuracy of data must be maintained. The Data Protection Act 1998 requires the data to be accurate and to be kept up to date. In the plan, develop appropriate maintenance procedures to sustain those records, where there is a requirement to preserve without further amendment or updating. This will ensure the accuracy of the data can be established and assured.

3.3 Effective maintenance policies and procedures for electronic record collections will help departments meet their obligations under FoI legislation. The Lord Chancellor's Code of Practice on the Management of Records under Freedom of Information, (version 21a, 21 June 2000) provides the following guidance.

- *Records of a business activity should be complete and accurate enough to allow employees and their successors to undertake appropriate actions in the context of their responsibilities, to facilitate an audit or examination of the business by anyone so authorised, protect the legal and other rights of the authority, its clients and any other person affected by its actions, and provide authenticity of the records so that the evidence derived from them is shown to be credible and authoritative.*
- *A tracking system should be used to control the movement and location of records. This should be sufficient to ensure that a record can be easily retrieved at any time, that any outstanding issues can be dealt with, and that there is an auditable trail of record transactions.*

BENEFITS: WHY SUSTAINABILITY IS IMPORTANT

- *A contingency or business recovery plan should be in place to provide protection for records which are vital to the continued functioning of the authority.*

Wherever possible, co-ordinate work to meet the Code of Practice with that to meet the requirements of the 2004 programme.

4 Key elements for a 2004 compliant sustainability plan

Rationale and essential prerequisites

4.1 Electronic records that organisations require to be retained for business purposes need to be brought into a managed environment to sustain access for as long as they are required. This requires a plan, which will affirm the principles and performance measures against which any sustainability initiative can be assessed. This section identifies the issues that need to be addressed in sustainability plan. Most organisations will have established some form of business contingency or continuity plan (also known as a disaster recovery plan). Many of the mechanisms required for a compliant plan should in part have been addressed by such documents. However they will need to be revised to take account of the context of the 2004 target if the plan is to be effective.

4.2 In order to develop the policies, procedures and mechanisms to achieve the Modernising Government target of storing and retrieving records electronically by 2004 all organisations need to develop a plan which:

- *has the formal approval and support of management*
- *is linked to the organisation's 2004 strategic planning process*
- *identifies management roles and allocates them to people or units*
- *includes appropriate resources costings and allocations*
- *is co-ordinated with the organisation's business continuity and IT security management policies and procedures*
- *establishes review mechanisms to ensure the plan is implemented effectively and is kept up-to-date and relevant.*

Initial high level plan

4.3 In order to meet the September 2001 milestone departments must develop an initial high level plan.. More detailed implementation plans can then be developed later to a time scale identified in the high level plan. In order to be effective the plan needs to reference the issues listed at 4.4, the completion criteria provided at 4.8 and the quality indicators at 4.10.

Issues

4.4 It is necessary to have established an inventory of your digital records before you start planning for their preservation. If you do not know the extent of the record preservation problem, the type of records which need to be preserved, or who and what is creating those records, you cannot hope to make an effective preservation plan. You will need to refer to the inventory to produce estimates of quantities to be preserved now and in the near future. The inventory must also indicate what types of software are producing the records and in what systems. Guidance on creating inventories is provided in the toolkit *Guidance for an inventory of electronic record collections*.

4.5 Following a risk assessment an incremental approach may be the most effective approach to ensure an organisation can focus on those records, which are deemed to be either of the highest value or the greatest risk of loss. This will also provide scope for research for the remainder particularly where complex technical issues are involved. However, it should be noted that electronic records also offer the potential for reuse, such as using one report as the basis for a later one. Caution should be taken before extrapolating from existing patterns of use

4.6 As well as having an inventory, you will have to make some decisions on retention and disposal of the records in it, in order to implement the preservation plan. Records whose retention period is short may require no special action, whereas those which must be preserved for longer periods, or permanently, clearly require some intervention for their preservation and sustainability to be assured.

4.7 Each organisation will have unique issues to address; however, certain ones will be common. The high level plan must take account of the following:

- communicate to departmental providers the need to maintain designated non-active electronic records
- define mechanisms to ensure media refreshment/migration at appropriate intervals
- ensure access to records on or generated by legacy systems is maintained when organisations change their technological platforms
- identify the full costs of ensuring continued access to required records.

Criteria for an effective plan

4.8 An effective plan has to define continuing access requirements: the known hardware and software requirements; storage, backup and media refreshment policies; provide for review mechanisms to ensure appropriate migration paths are established. The high level plan must:

- identify categories, volumes and format of the records to be maintained, as listed in the inventory of record collections
- identify strategic mechanisms for how to ensure sustainability for each category
- support the continued management of record collections captured in departmental inventories (which have continuing business value) and new collections of records as they are created or identified.

Section 6 provides more information on the options available.

4.9 Having established this, determine mechanisms and resources for:

- identifying legacy systems and software formats
- establishing timescales for the evaluation of legacy systems and formats to determine the appropriate migration strategy

- developing proposals to establish or migrate to new systems which explicitly address any residual needs to maintain records on the old systems and platforms
- clarifying media refreshment/migration policies to ensure that required records and relevant metadata held on obsolescent formats are written to appropriate new media in a controlled manner
- clarifying control and management of back-up copies to prevent them frustrating compliance with official retention schedules and data protection requirements
- periodic review of the software formats required to access records to determine future requirements for software viewers and/or software migration
- ensuring that the plan is consistent with requirements of the e-Government Interoperability Framework (e-GIF)
- identifying and safe-guarding all essential documentation (system management and user manuals and other technical documentation) needed to support access to records on legacy platforms.

Quality indicators to assess effectiveness of the plan

4.10 The following quality indicators should be used to assess the efficacy of the products required by the strategic plan:

- outcomes are linked to implementation and migration planning for ERM systems – this should be consistent with the products and timescales in corporate strategic plans for ERM
- specific provision is made within business contingency planning to ensure continued access to records held on legacy formats
- mechanisms are provided to review at specified intervals software formats supported by the organisation, to determine migration strategies and requirements prior to obsolescence
- the plan supports and endorses a variety of approaches for both software migration and use of viewers/emulators to ensure continued access.

Prioritising the work – use of plans and timetables

4.11 The high level plan should assign responsibilities for roles and activities. Detailed plans will need to include a timetable that clearly assigns specific responsibility for all activities, and estimates the schedule for undertaking the assessment of the technical requirement for each electronic record collection. The scope and complexity of this plan will vary from one organisation to another. Where required each task should be clearly explained with an estimate of the effort required and the completion deadline. The plan should also include a schedule of which departments and units will be consulted, by whom, and when.

4.12 The products of the plan should be clearly defined along with appropriate quality measures.

4.13 A checklist of the required actions together with time-scales and the identity of the person(s) responsible is a useful planning tool and it will also help in the monitoring of progress against the overall target.

Communication strategy

4.14 Where a need for additional information is identified it will be necessary to communicate to all the data owners, custodians and users of these electronic record collections that their help will be needed. Where necessary, senior management support should be sought to ensure compliance, in the context of progress towards the 2004 target.

4.15 The potential benefits for ordinary users should also be identified and promulgated. This will help make the process a valued and meaningful exercise to the staff and managers consulted, during the process, by emphasising the internal benefits that will accrue. These include:

- sustained accessibility to corporate information
- better information and knowledge management
- assured maintenance of the required records
- more efficient and effective use of the available storage capacity.

Management of relations with IT providers

4.16 Where organisations have outsourced their IT provision to an external provider it will be necessary to explore what the impact of a sustainability maintenance plan will have on the contract. The need for sustainability will need to be clarified to ensure the supplier is aware of the organisation's needs as the archiving of data has traditionally been handled by IT personnel but the attributes of this process have very little to do with archiving as it is traditionally understood in the records management community. Usually data is copied wholesale on to duplicate media and held for a prescribed time, to be deleted after a specific interval in order to release the media for re-use with new data (not necessarily because the business need for these records had actually expired).

4.17 It is possible that little or no provision for sustainability was made in the initial contract and it will be necessary to make them aware of the departmental requirement in future negotiations.

Business contingency/continuity planning

4.18 As stated at the beginning of this section most organisations will have established some form of business contingency or continuity plan (also known as a disaster recovery plan). Many of the mechanisms required for a compliant plan should in part have been addressed by such documents. However a contingency plan is not an adequate substitute for a sustainability plan and they will need to be revised to take account of the context of the 2004 target if the plan is to be effective.

Legal admissibility and evidential weight

4.19 *The BSi Code of Practice for Legal Admissibility and Evidential Weight of Information Stored Electronically DISC PD 0008* together with its accompanying *Compliance Workbook PD 0009* defines current guidance on this topic. The sustainability plan should reference both documents and make provision so that all solutions to maintain records are benchmarked against them. It is essential that the organisation is aware of the value of information that it stores, and executes its responsibility in accordance with its provisions. Particular care should be given before any migration strategy is approved. The Code of Practice provides the following guidance on migration:

- *There should be provision for migrating data files, including metadata, index data and audit trails, to new technology without loss of integrity, and with sufficient migration process documentation to allow the integrity of the data to be established beyond any reasonable doubt at any time in the future.*

5 Access

5.1 This section identifies the various access needs that an effective sustainability plan should provide for. Records maintained without adequate access provision are effectively worthless. The critical elements can be summarised as follows:

- identify who can make requests, and who can execute them
- understand management parallels with paper records
- beware of using past access to predict future access patterns
- take disaster recovery into account in planning

5.2 Access must be capable of being defined in response to an organisation's business needs and is likely to vary according to the organisation's information requirements. Records may need to be made available to the entire organisation or to a designated part of the organisation; there may be a series of levels or views depending on user access permissions. As time proceeds, ultimately the chief end user will be the DRO and the records management staff, who will require access for appraisal and final disposition of the records.

5.3 The Freedom of Information Act will also define a new set of access requirements which will require appropriate access regimes and the need to develop departmental FOI publications will give rise to a new set of resource requirements.

5.4 In addition, an organisation may wish to maintain electronic documents because the information they contain can be more readily accessed and exploited to the benefit of the business than when it is held on paper in a registered file located in a Registry.

5.5 The nature of the various preservation solutions chosen, and their cost, will be affected by the access needs to those records stored within it.

5.6 This situation is comparable to that for conventional records. Those which need to be accessed frequently are kept near to hand, and typically cost more to store per unit volume, but have low cost penalties for access. Those which are not expected to be required frequently, if at all, may be stored offsite in a manner which affords cost benefits for storage, but carries penalties in time and money on the rare occasions when access is required.

5.7 It must be recognised that future access patterns cannot be predicted with certainty. You need to make an assessment of the risks should your predictions be incorrect.

Who needs access

5.8 The first task is to identify the requirements for access to the records. This question extends to both who may request the records, and who will actually have the means of access to carry out the request. The analogy with paper records is straightforward: many people may request files, but only certain people have direct access to the file repository to retrieve them.

5.9 There is also a need to identify those who will need management access, in order to manipulate record metadata, remove records whose continued retention is no longer necessary, or add new material or metadata.

How access is provided

5.10 In environments, using a shared file store or some form of EDRM, it is normal practice to provide a means for those who require records to retrieve them directly from the system, which stores and maintains the records. This does not mean providing direct, on-line access to the stored records, but providing read-only access via a computer network which allows records to be retrieved as and when they are needed. This approach is typically popular with those needing the records and reduces the burden on records management staff. It will place greater demands on network resources. Whatever method is selected will need to take account of the estimates for frequency of access.

Frequency of access

5.11 The need to access has also to take account of the anticipated reasons and frequency. The reason for retention, and the likely frequency of use can determine the mode of storage and retrieval. Essentially there are four types of storage. These are:

- On-line – (for example, on servers on magnetic disk)
- Near-line (for example, CD juke-box or robotic tape storage)
- Far-line (for example, an on-line index with the records held on disk or tape in off-line storage)
- Off-line (for example, index and objects stored on media externally to the network but physical retrieval possible upon request).

In practice the business requirement will mean the first two options are preferred for their greater convenience. However, the economics of far-line and off-line storage may be very attractive if the use of the records is estimated as being residual and very infrequent.

5.12 It can be difficult to predict how often access will be required to records. In many cases installing systems, which makes it easier to access older material, may increase the frequency with which it is accessed. Electronic records also offer the potential for reuse, such as using one report as the basis for a later one. Caution should be taken before extrapolating from existing patterns of use.

5.13 However, the goal at this stage is get an estimate at the level of an order of magnitude. What are the anticipated types of access - that is, are specific documents or database records

required, or is access likely to involve an entire collection of documents, or an entire year's worth of database entries? The goal is to establish resource requirements generated by access requests.

5.14 Access requirements can be characterized by the following the following estimates:

- total number of retrieval requests in a given period
- average number of documents requested
- average total size of request in megabytes
- for databases, cost of database query (rows retrieved or examined)
- anticipated methods of retrieval (e.g use of keywords, full text indexes and thesauri).

Disaster recovery

5.15 An effective disaster recovery plan should be coordinated with the departmental risk register – as failure to ensure effective management of an identified risk as senior staff may be personally liable.

5.16 Access requirements specific to disaster recovery are important to identify. They will be used to shape the methods you use for storing contingency copies of your digital records.

5.17 Departments will identify, when they evaluate the contents of their inventories, which records fall into the class of vital and emergency records. It is also necessary to identify what actions may already be being taken to ensure continued access to digital records in the event of a disaster. For instance, regular backups of record-keeping systems may be created, and copies stored in an off-site location. It is also necessary to determine what role the electronic records repository will be required to play in the event of a disaster elsewhere in the organisation:

- will the archived version be required to provide records to replace copies lost or destroyed elsewhere?
- how rapidly will those copies be required?
- how will they be provided - via a network, tapes, or some other method?

5.18 All IT departments and providers should have established a disaster recovery plan and this should be explicitly extended to records being maintained for business purposes. 'Disasters' for IT systems can mean something far less drastic than fire or flood. Failure of a disk, bugs in software, or simple user error can have catastrophic effects. It is however easy to protect records against most of these events. Creating extra copies on tape or disk of all preserved records, and storing that copy elsewhere, will protect against many events. The following points need to be addressed:

- how often to create these additional backups?
- how often will they be checked for readability?
- how long it will take to recreate the system from these backups should this be necessary?

It is essential to establish the time-to-recovery for emergency records. Periodic tests should be conducted to confirm that recovery procedures actually work.

Vital and emergency records

5.19 Any preservation plan needs to distinguish between vital records and emergency records and to make appropriate provision for their survival. Vital records are those records whose long-term preservation must be ensured to allow the continuation of organisational functions. In the event of a disaster, an organisation may not need immediate access to vital records, but it must be confident of their safety. Emergency records are those records to which rapid access is necessary following a disaster; records without which your organisation cannot function for more than a few days. It should be understood that not all emergency records are vital. Many vital records are not emergency records.

5.20 An example of vital records would be financial records, which are required for auditing purposes only. One can get by without these for some months, but eventually the auditor will want them. If there are delays getting hold of them because of some disaster, the auditor will understand, but you cannot manage without them forever as the accounts are required to be audited. Documents necessary to prove title to something, or original contracts, also fall into this category. Such records are vital (i.e. they must be sustained) but they are not essential to the management of the emergency.

5.21 Emergency records are those, which you need immediate access to but which are not critical to the business in the long term. Current contact directories, telephone lists and procedural manuals would fall into this category. Out of date copies of internal directories and manuals are rarely relevant in an emergency but the absence of current documentation could make timely recovery very difficult and could compromise the quality of the restored service.

Checklist

Information category	Comments
Who needs access	Identify who needs access to what; can access be direct, or must it be via records management staff?
Frequency of access	How often is access to preserved records likely to occur? When it does happen, will whole batches of records be required, or just individual items?
Urgency of access	When access is needed, what are acceptable delays? What are the costs of increased delays? What is the maximum acceptable delay?
Disaster recovery access	What role will your record repository play in disaster recovery? Conversely, how quickly should your record repository recover from its own disasters? What are the vital records?

6 Options

Summary

6.1 This section identifies four options, mapped on the matrix below. Follow the option in a quadrant which is best for records and documents where:

	Maintain on existing platform	Migrate to an ERM system
Native format	Access is required for not more than 2-3 years and No significant need to associate with subsequent records for business continuity/accountability for retrieval purposes exists	Access is required for up to 3 years and Significant need to store and retrieve with future created records for business continuity/accountability exists
Standard format	Access is required for more than 2- 3 years and No significant need to associate with subsequent records for business continuity/accountability for retrieval purposes exists	Access is required for more than 3 years and Significant need to store and retrieve with future created records for business continuity/accountability exists

6.2 Duration of the sustainability period will vary according to departmental needs. In most departments this is unlikely to exceed 30 years and will in many cases be far shorter. However, certain departments and agencies will need to hold records for periods in excess of 30 years. It is likely that increased dependence on electronic record-keeping will generate a requirement for organisations to retain data for longer periods, as already happens with paper based records; a relevant example are personal staff files held by personnel or human resource departments which are kept for many decades. In such cases the long-term preservation concerns are similar to those of a specialist archive such as the Public Record Office, which aims to preserve selected records indefinitely. It is anticipated that as the specifications for long-term storage are fully defined the PRO will develop additional guidance.

6.3 This section explores the various options available to departments in maintaining existing electronic records and documents during the transition period to 2004 and while longer term strategies are being developed. Clearly where records need only to be retained for a brief period (for example, two to three years) maintenance of the records on their original systems will remain an attractive and viable proposition. However if the material is already on an obsolescent platform action may be required for example software migration or adoption of emulators.

6.4 Consideration of business objectives is important when deciding which approach will be most appropriate. Records managers must note that this remains an organisation's own responsibility and each one must develop appropriate working practices and procedures to care for public records within the context of their own business.

Storage of documents in native format on existing platforms

6.5 Where the original application is to be used to access the records, it is imperative that the departmental IT strategy makes provision for an annual review of all applications and platforms to ensure that appropriate support is given for all originating technology. It is also necessary to ensure that any policy decision to change to another platform takes account of both the migration requirements of records held in native formats and the requirement to preserve the integrity of the information. It should do this by storing the records together with the contextual metadata in a tightly bound relationship in a stable area of the organisation's workspace to ensure they cannot be modified or deleted by users.

6.6 For many departments the initial solution is to continue to maintain electronic records in the current software format on the existing hardware platform. This will probably be sufficient for records which need to be sustained for a period of up to two years providing the organisation can continue to maintain the relevant hardware and continued access to the relevant versions of the software are also guaranteed. The following issues need to be explicitly addressed:

- what version of the software are the records held in?
- is the software actively supported?
- does the organisation possess the appropriate licences to continue to access records in that format?
- are the records dependent on specific hardware?

6.7 Storage in native format on existing platforms can only be a short-term solution and if the records are to remain accessible beyond two to three years other options will have to be considered. These include:

- are appropriate viewers available to end dependence on the authoring software?
- do the viewers provide adequate access to the record – (some viewers do not support access to information created by macros)?
- can the records be rendered into an interoperable format (for example, an e-GIF approved format)?
- an annual review of software formats used by the organisation to determine if a format is becoming obsolescent

6.8 If the records are still required for business purposes it is critical they do not remain in a format which cannot be accessed by those authorised to view the records. Effective migration

or emulation strategies to avoid dependence on the old software are needed. The next section explains the implications of these two approaches.

Migration strategies and emulation environments

6.9 This section discusses the relative merits of migration and emulation strategies. It is unlikely that the strategies developed to meet the September 2001 milestone will be able to determine longer term preservation strategies at this stage – in most cases government organisations will want to maintain their options. This discussion does, however, provide a context for taking shorter-term decisions.

6.10 Longer term preservation in an in-house Electronic Document and Record Management (EDRM) system, or a dedicated archival resource system will require some form of periodic action to ensure the information will remain accessible as the technological environment changes. This typically involves one of two techniques: migration or emulation. The tools available, the cost of deployment, and the type of information being preserved will influence which is chosen. It is possible to use both techniques within the same system, each applying to a specific class of information.

Migration denotes the periodic transformation of the data from one format to another or from an earlier version to a current version of the same format. examples of these would be the conversion of a document from Microsoft *Word* format to PDF or from *Word 6* to *Word 7*.

Emulation means keeping the original file formats, but producing a software application or complete system that allows original software to be used to access the information.

6.11 At the present time, migration is by far the most common strategy employed in any digital preservation system. Emulation is still primarily a topic of research, but is being employed to good effect in certain specialised contexts. For instance, emulation systems are available for current PCs running Microsoft Windows allowing them to run software and access data produced for early microcomputers such as the BBC Micro and Sinclair Spectrum.

6.12 Both strategies are distinct from any mechanisms you may use to make information accessible at any particular time which require transformation of the preserved files to some other form. For instance, you may have a quantity of documents in WordPerfect format, which you choose to preserve in this native format. To make access easier, you may provide copies from the archive in a more modern format, such as Microsoft Word or PDF. In doing so, you have not altered the format of the preserved files, and thus neither emulation nor migration has taken place.

6.13 In order to plan for future migration, you must ensure that your files are currently stored in a format, which is well-defined, preferably via a public standard of some form. The e-GIF framework document lists many appropriate formats and is available at www.govtalk.gov.uk

6.14 If files are stored in a proprietary format, future migration may prove difficult if the original supplier is no longer in business or no longer supports the file format in question. It is difficult to know how often migration will be necessary, as depending on the format the frequency will vary. Some office applications may need to be migrated after 3 years whereas in the case of some open or non proprietary formats once every 10 to 15 years is reasonable. Periodic review of storage formats should allow you to plan any migration at least one year in advance.

6.15 The advantages of migration are that it makes no special demands on the computer systems of those who must access the records – all conversion is handled within the preservation environment. The primary disadvantage is that migration is often imperfect, in that some attributes of the record will be lost or altered. Ensuring that key metadata is stored separately in the system (i.e. not just embedded in the file) alleviates some concerns. Any proposal to migrate needs to be informed about the likely loss of functionality involved in the migration in order to permit a risk analysis to be undertaken before a decision is taken to proceed.

6.16 It is important to ensure that someone with suitable technical understanding of the process involved tests the migration process thoroughly and that the process is subject to independent audit. Where some information is unavoidable lost or altered by the process, ensure this is clearly documented in any catalogue of the records.

6.17 You can defend against some of the disadvantages of migration by preserving the original files as well as the migrated files. The costs of this are likely to be less than might be thought, since future storage will cost a fraction of what it does today. For example 10 Gigabytes of disk in 2001 costs 2% of the 1996 cost. The advantage of this policy is that you can assume that the migrated files will suffice for the majority of accesses to your archive.

6.18 Emulation requires providing appropriate emulation software on every system, which might need access to preserved files. This can prove extremely expensive. Alternatively, if only one copy of the emulation software is held within the archive, it requires people to physically travel to the archive in order to access the files within it. This is likely to prove unpopular and inconvenient.

6.19 Emulation can operate at multiple levels. At the most complex, an entire system is emulated at the hardware level. To utilise such an emulator, one must preserve not only the original files, but also the application to read them and the operating system which the application needs. The advantages of such an approach are that the one emulator can be used to provide access to any file generated on a system of this type, provided that the applications have also been preserved. The primary disadvantage is that, as time moves on, the interface provided by the original applications will be increasingly unfamiliar to current researchers or staff. For example, computer systems of the 1960s typically took input in punched card or paper tape and provided output on large-format paper or more cards or tape. They can be emulated successfully today, but using such an emulation requires knowledge of the arcane control languages of the time, a skill which most do not have and do not wish to acquire.

6.20 Emulation of applications is an intermediate route and may prove more sustainable. This involves creating a single emulation application for each particular file format and each host system type. The emulation runs on current computer systems, but reads data from older ones without the need for conversion. Again, a single emulation package can provide access to all files of a particular type. Microsoft *Excel*, which can read spreadsheets created by older spreadsheet systems such as Lotus 1-2-3, is an example of a simple emulator but any editing would require use of the editing tools in Excel and if the changes were saved the document would effectively be migrated to *Excel*.

6.21 Application emulation has two drawbacks. One is that it effectively uses the same techniques as migration, and so similar risks may exist that not all attributes of the original file will be preserved. The other is that it does not preserve the same interface to the use of the file, and this may be relevant in some contexts. For an archive with many file formats in it, application emulation also requires the use of many emulators, which will present both cost and maintenance problems.

Before considering emulation as a strategy, ask yourself the following:

- Am I confident an emulator will exist or can be developed when it is needed?
- Will the cost of the emulator prohibit effective use of the stored records?
- Will those needing access to the records need training in the use of the emulator, and what will be the cost of that?
- Do I have an alternative to the use of emulation?

Migration to standardised formats on existing platforms

6.22 The PRO recommends that where records are required to be maintained beyond three years a strategy is developed and implemented which should provide mechanisms to preserve data and metadata in a format that is independent of the particular hardware and software used to create them; in this way the integrity of the records should be preserved along with the data content.

6.23 Migration policies should always aim to avoid dependence on proprietorial formats such as Microsoft *Word* or Lotus 1-2-3 by rendering records held in their original native formats to open interoperable formats such as those listed in the e-GIF document. However it is essential that the most appropriate format needed to support the residual business need is selected. There is no advantage for example in adopting a policy that all electronic documents are rendered to a static image format such as TIFF or GIF if users need to be able to search the text and copy and paste text into new documents. Migration must always try to support the business need as well as preserve the information content.

6.24 Extensible Mark-up Language (normally referred to as XML) is widely viewed as one of the preferred interoperable formats for both data and metadata. Its flexibility and use of browser technology are amongst some of its obvious assets. This toolkit recommends that XML should be seriously considered when any migration strategy is being developed or

revised. However its true potential for UK government will become apparent as and when XML schemas for UK government are formalised.

6.25 Operating procedures will need to address the following:

- compliance with *BSI/PD0008* and proof of compliance
- migration policies for hardware and systems and application software, including creation of audit trails establishing successful migration with no loss of data
- hardware refresh
- maintenance of access
- checking of back-ups to prove continued utility and availability.

6.26 Procedures must be implemented so that all necessary tasks are performed to ensure the successful, long-term preservation of electronic records. In addition, the roles needed to carry out these procedures must be defined and built into the job descriptions of members of staff. This latter is to ensure that not only are the procedures in place but that someone is tasked with doing them. The commitment of the senior management and budget holders is required to make available the resources needed to maintain preservation. These procedures will be long-term and will cover the entire life of the records.

6.27 All activities undertaken as a result of implementing these procedures must be logged on an audit trail and these records should comply with the recommendations contained in *BSI/PD0008*.

Storage in native format on managed EDRMS environments

6.28 As departments move to adopt electronic document and record management (EDRM) systems there are advantages to importing existing collections into the new environment as this will impose a corporate information structure, with appropriate access controls and audit trails. Plan to transfer records and documents into a new EDRM system where this is justified for reasons of business continuity or accountability.

6.29 However, importation into such an environment does not address the continued viability of the records if they remain in the format in which they were created. Migration, emulation or dependence on viewers will still be necessary. Storage of records in their native format on an EDRMS will still require active management and the issues outlined in 6.9 to 6.27 are equally relevant here. The PRO is actively pursuing work in this area and will make recommendations later in 2002.

Storage in standardised format on managed EDRMS environments

6.30 Some EDRMs offer the possibility of rendering to selected formats such as PDF. Other EDRMs do not feature this as part of the core functionality but they will support application of metadata to copies rendered into another format. All EDRMS will support importation of records into the managed store.

6.31 The main advantage of an EDRMs is that it will provide scope for additional metadata to be added regarding the management and status of a record or group of records. This then is supported by a full audit trail, which will document what actions were undertaken. The guidance in given in paragraphs 6.22 to 6.27 is also applicable when determining a migration path.

Best practice management of databases on existing platforms

6.32 Dealing with the preservation of existing databases requires some steps that are unique to this type of system. You will need to:

- classify the type of database
- determine which database components require capture
- in some cases, determine the frequency of capture
- determine when separate preservation measures will be required
- ensure capture of relevant technical and contextual metadata

Following this, you will need to ensure that your decisions are implemented and that they continue to be implemented.

Classifying a database

6.33 This task is a straightforward one and requires no technical knowledge of how the database operates. It does require knowledge of what information goes into the system and what can be retrieved. You need to determine two attributes of the database, each of which can take two possible values. This leads to the database falling into one of four classes. The attributes are summarised here and explained further overleaf.

Attributes of data permanence

Dynamic: A system in which old information is overwritten by new information.

Static: A system in which information, once added, is never removed nor overwritten.

Attributes of system lifetime

Open: A system to which information is continually added, where no clear beginning or end to the information is defined. Open systems can become closed when they are superseded or a decision is taken to decommission them.

Closed: A system designed to record a specific set of information, after which no more will be added. Most surveys fall into this category. Alternatively, a system no longer in use to which no more information can or will be added.

Determining what to capture

6.34 Having classified the database, you must consider what information in it requires preservation, and use the classification to determine how this will be achieved.

6.35 **Static, closed** databases do not offer any scope for decision as to how they will be preserved. By definition, they are systems that hold all information that was ever entered into them, and to which no new information will ever be added in future. Therefore, the only decision is whether they require preservation. If the systems are in active use, preservation within the system is adequate. Once the system has been retired, active steps are required to retrieve information from the system or its backups

6.36 **Static, open** databases are also straightforward to deal with. By definition, information is never overwritten in these systems, but is still being added to. In some cases, performance or operational requirements will mean that redundant information is periodically removed from such systems. One strategy is thus to ensure that information to be expunged is captured for preservation elsewhere. Another, if no expunging takes place, is simply to wait for the system to become defunct (hence static and closed) and deal with it then. Public access requirements may mean that you wish instead to actively extract information from the system on a periodic basis – annually or five-yearly – and pass this to the PRO for preservation.

6.37 **Dynamic, closed systems** do not offer scope for decision on their preservation, although they are often a demonstration of inadequate preservation decisions in the past. In a system, which is both dynamic and closed, we know that information has been entered into the system and subsequently destroyed or amended before we could capture it. Because the system is now described as closed, the only possibility is that it is no longer in active use (although information may still be being retrieved from it, no new information is being entered.) Treat as for static, closed systems but ensure that the past loss of information is understood and documented by those responsible for cataloguing the information.

6.38 **Dynamic, open systems** present the greatest challenges and the greatest variety of potential preservation scenarios. It is likely that following evaluation of the content of such

systems that these systems will have a higher priority when preservation is considered. Departments should take this into account when determining their sustainability strategy. A number of approaches are possible.

- The snapshot – this means that a copy of all data in the system is preserved at some instant in time. Snapshots are usually taken periodically, and will need to be converted to a standard form rather than kept in its native form. Snapshots are a simple method to apply, but offer a poor picture of the information flow through a system. In a casework system, for example, a snapshot tells us how many outstanding cases existed at the time of the snapshot, but not how many cases were dealt with per month or per year. A snapshot may not necessarily capture all relevant information – for example, coding systems – if part of this is held in a physically separate system. It should be noted that snapshots are likely to lose the interactivity of the system, and some of the presentational context, if such functionality was supplied by applications such as Java or Active X.
- The audit trail – by capturing a single snapshot, and then capturing information about every alteration made to records in the system, you can effectively convert a dynamic, open system to a static, open system. The audit trail is only ever appended to, never overwritten. If this option is taken, you must ensure that you also have full documentation about the format of the audit trail and can use it to reconstruct records within the system. You must also ensure that all relevant changes are audited: in some systems only some changes may create audit-trail records.

Records no longer required for business purposes

6.39 Electronic documents which are considered to be the primary record may ultimately be selected for transfer to the Public Record Office if they originate from an organisation subject to public records legislation. Electronic record collections which are no longer required by the department should, in all cases, be:

- brought to the attention of the relevant PRO Client Manager
- submitted for appraisal by the normal mechanisms
- retained until a confirmed decision is taken on final disposal

6.40 Records no longer required by the department and not selected for transfer to the Public Record Office should be deleted. However, a record of this decision and the subsequent destruction action should be maintained thereafter.